```
         RAW SEQUENCE LISTING              DATE: 12/07/2000
         PATENT APPLICATION:  US/09/464,377    TIME: 10:03:40

         Input Set : A:\726sl.txt
         Output Set: N:\CRF3\12072000\I464377.raw

 4  <110> APPLICANT: University of Southern California
 5         Stallcup, Michael R.
 6         Chen, Dagang
 7         Hong, Heng
 8         Asward, Dana W.
10  <120> TITLE OF INVENTION: REGULATION OF GENE EXPRESSION BY PROTEIN
11         METHYLATION
13  <130> FILE REFERENCE: 13761-726
15  <140> CURRENT APPLICATION NUMBER: US 09/464,377
16  <141> CURRENT FILING DATE: 1999-12-15
18  <150> PRIOR APPLICATION NUMBER: US 60/112,523
19  <151> PRIOR FILING DATE: 1998-12-15
21  <160> NUMBER OF SEQ ID NOS: 10
23  <170> SOFTWARE: FastSEQ for Windows Version 4.0
25  <210> SEQ ID NO: 1
26  <211> LENGTH: 3124
27  <212> TYPE: DNA
28  <213> ORGANISM: Mus musculuc
30  <400> SEQUENCE: 1
31  agggggcctg gagccggacc taagatggca gcggcggcag cgacggcggt ggggccgggt     60
32  gcggggagcg ctggggtggc gggccggggc ggcgcggggc cctgcgctac agtgtctgtg    120
33  ttcccgggcg cccgcctcct cactatcggc gacgcgaacg gcgagatcca gcggcacgcg    180
34  gagcagcagg cgctgcgcct tgaggtgcgc gccggaccag acgcggcggg catcgccctc    240
35  tacagccatg aagatgtgtg tgttttcaag tgctcggtgt cccgagagac agagtgcagt    300
36  cgtgtgggca gacagtcctt catcatcacc ctgggctgca acagcgtcct catccagttt    360
37  gccacacccc acgatttctg ttctttctac aacatcctga aaacctgtcg gggccacaca    420
38  ctggagcgct ctgtgttcag tgagcggaca gaggaatcct cagctgtgca gtacttccag    480
39  ttctatggct acctatccca gcagcagaac atgatgcagg actatgtgcg gacaggcacc    540
40  taccagcgtg cgatcctgca gaaccacacg gacttcaagg acaagatcgt tctagatgtg    600
41  ggctgtggct ctgggatcct gtcatttttt gctgctcaag caggagccag gaaaatttat    660
42  gcagtggaag ccagcaccat ggctcagcat gcagaggtcc tggtgaagag taacaatctg    720
43  acagaccgca tcgtggtcat ccctggcaaa gtagaggagg tctcattgcc tgagcaagtg    780
44  gacattatca tctcagagcc catgggctac atgctcttca atgaacgaat gctcgagagc    840
45  tacctccatg ccaaaaagta cctgaagcct agtggaaaca tgttccccac cattggtgat    900
46  gtccacctcg cacccttcac tgatgaacag ctctacatgg agcagttcac caaagccaac    960
47  ttccggtacc agccatcctt ccatggagtg gacctgtcgg ccctcagagg tgccgctgtg   1020
48  gatgagtact tccggcaacc tgtggtggac acatttgaca tccggatcct gatggccaaa   1080
49  tctgtcaagt acacagtgaa cttcttagaa gccaaagaag gcgatttgca caggatagaa   1140
50  atcccattca aattccacat gctgcattca gggctagtcc atggcttggc cttctggttc   1200
51  gatgttgctt tcattggctc cataatgacc gtgtggctat ccacagcccc aacagagccc   1260
52  ctgacccact ggtaccaggt ccggtgcctc ttccagtcac cgttgttttgc caaggccggg   1320
53  gacacgctct cagggacatg tctgcttatt gccaacaaaa gacagagcta tgacatcagt   1380
54  attgtggcac aggtggacca gacaggctcc aagtccagta acctgctgga tctaaagaac   1440
55  cccttcttca ggtacacagg tacaacccca tcacccccac ctggctcaca ctacacgtct   1500
56  ccctcggaga atatgtggaa cacaggaagc acctataatc tcagcagcgg ggtggctgtg   1560
57  gctggaatgc ctactgccta cgacctgagc agtgttattg ccggcggctc cagtgtgggt   1620
58  cacaacaacc tgattccctt agctaacaca gggattgtca atcacaccca ctcccggatg   1680
```

Input Set : A:\726sl.txt
Output Set: N:\CRF3\12072000\I464377.raw

```
59 ggctccataa tgagcacggg cattgtccaa ggctcctcag gtgcccaggg aggcggcggt    1740
60 agctccagtg cccactatgc agtcaacaac cagttcacca tgggtggccc tgccatctct    1800
61 atggcctcgc ccatgtccat cccgaccaac accatgcact atgggagtta ggtgcctcca    1860
62 gccgcgacag cactgcgcac tgacagcacc aggaaaccaa atcaagtcca ggcccggcac    1920
63 agccagtggc tgttcccct tgttctggag aagttgttga acacccggtc acagcctcct    1980
64 tgctatggga acttggacaa tttgtacac gatgtcgccg ctgccctcaa gtaccccag    2040
65 cccaaccttt ggtcccgagc gcgtgttgct gccatacttt acatgagatc ctgttggggc    2100
66 agccctcatc ctgttctgta ctctccactc tgacctggct ttgacatctg ctggaagagg    2160
67 caagtcctcc cccaaccccc acagctgcac ctgaccaggc aggaggaggc cagcagctgc    2220
68 caccacagac ctggcagcac ccaccccaca acccgtcctt gcacctcccc tcacctgggg    2280
69 tggcagcaca gccagctgga cctctccttc aactaccagg ccacatggtc accatgggcg    2340
70 tgacatgctg ctttttttaa ttttattttt ttacgaaaag aaccagtgtc aacccacaga    2400
71 ccctctgaga aacccggctg gcgcgccaag ccagccagta ctgttcctag gcccagaagg    2460
72 tctaggtgag gggtggccct gtcaagcctt cagagtgggc acagccccte ccaccaaagg    2520
73 gttcacctca aacttgaatg tacaaaccac ccagctgtcc aaaggcctag tccctactct    2580
74 ctgctactgt cctgtcctga gccctgaagg ccccccctcca tcaaaagctt gaacaggcag    2640
75 cccagagtgt gtcaccctgg gctactgggg cagacaagaa acctcaaaga tctgtcacac    2700
76 acacacaagg aaggcgtcct ctcctgatag ctgacatagg cctgtgtgtt gcgttcacat    2760
77 tcatgttcta cttaatcctc tcaagacagc aacccgtggca aggagcctc cagggacctc    2820
78 cccagacaag aagaaaagca aacaaggaag ggtgattaat aagcacaggc agtttcccct    2880
79 attcccttac cctagagtcc ccacctgaat ggccacagcc tgccacagga accccttggc    2940
80 aaaggctgga gctgctctgt gccaccctcc tgacctgtca gggaatcaca gggccctcag    3000
81 gcagctggga accaggctct ctcctgtcca tcagtaatac tccttgctcg gatggccctc    3060
82 ccccaccttt atataaattc tctggatcac ctttgcatag aaaataaaag tgtttgcttt    3120
83 gtaa                                                                   3124
85 <210> SEQ ID NO: 2
86 <211> LENGTH: 608
87 <212> TYPE: PRT
88 <213> ORGANISM: Artificial Sequence
90 <220> FEATURE:
91 <223> OTHER INFORMATION: Deduced amino acid sequence of CARM1
93 <400> SEQUENCE: 2
94 Met Ala Ala Ala Ala Ala Thr Ala Val Gly Pro Gly Ala Gly Ser Ala
95 1               5                   10                  15
96 Gly Val Ala Gly Pro Gly Gly Ala Gly Pro Cys Ala Thr Val Ser Val
97             20                  25                  30
98 Phe Pro Gly Ala Arg Leu Leu Thr Ile Gly Asp Ala Asn Gly Glu Ile
99          35                  40                  45
100 Gln Arg His Ala Glu Gln Gln Ala Leu Arg Leu Glu Val Arg Ala Gly
101          50                  55                  60
102 Pro Asp Ala Ala Gly Ile Ala Leu Tyr Ser His Glu Asp Val Cys Val
103 65                  70                  75                  80
104 Phe Lys Cys Ser Val Ser Arg Glu Thr Glu Cys Ser Arg Val Gly Arg
105                  85                  90                  95
106 Gln Ser Phe Ile Ile Thr Leu Gly Cys Asn Ser Val Leu Ile Gln Phe
107             100                 105                 110
108 Ala Thr Pro His Asp Phe Cys Ser Phe Tyr Asn Ile Leu Lys Thr Cys
109          115                 120                 125
110 Arg Gly His Thr Leu Glu Arg Ser Val Phe Ser Glu Arg Thr Glu Glu
```

RAW SEQUENCE LISTING   DATE: 12/07/2000
PATENT APPLICATION: US/09/464,377   TIME: 10:03:40

Input Set : A:\726sl.txt
Output Set: N:\CRF3\12072000\I464377.raw

```
111      130                     135                     140
112 Ser Ser Ala Val Gln Tyr Phe Gln Phe Tyr Gly Tyr Leu Ser Gln Gln
113 145                     150                     155                 160
114 Gln Asn Met Met Gln Asp Tyr Val Arg Thr Gly Thr Tyr Gln Arg Ala
115                     165                     170                 175
116 Ile Leu Gln Asn His Thr Asp Phe Lys Asp Lys Ile Val Leu Asp Val
117                 180                     185                     190
118 Gly Cys Gly Ser Gly Ile Leu Ser Phe Phe Ala Ala Gln Ala Gly Ala
119             195                     200                     205
120 Arg Lys Ile Tyr Ala Val Glu Ala Ser Thr Met Ala Gln His Ala Glu
121         210                     215                     220
122 Val Leu Val Lys Ser Asn Asn Leu Thr Asp Arg Ile Val Val Ile Pro
123 225                     230                     235                 240
124 Gly Lys Val Glu Glu Val Ser Leu Pro Glu Gln Val Asp Ile Ile Ile
125                     245                     250                     255
126 Ser Glu Pro Met Gly Tyr Met Leu Phe Asn Glu Arg Met Leu Glu Ser
127             260                     265                     270
128 Tyr Leu His Ala Lys Lys Tyr Leu Lys Pro Ser Gly Asn Met Phe Pro
129         275                     280                     285
130 Thr Ile Gly Asp Val His Leu Ala Pro Phe Thr Asp Glu Gln Leu Tyr
131         290                     295                     300
132 Met Glu Gln Phe Thr Lys Ala Asn Phe Arg Tyr Gln Pro Ser Phe His
133 305                     310                     315                 320
134 Gly Val Asp Leu Ser Ala Leu Arg Gly Ala Ala Val Asp Glu Tyr Phe
135                     325                     330                     335
136 Arg Gln Pro Val Val Asp Thr Phe Asp Ile Arg Ile Leu Met Ala Lys
137             340                     345                     350
138 Ser Val Lys Tyr Thr Val Asn Phe Leu Glu Ala Lys Glu Gly Asp Leu
139         355                     360                     365
140 His Arg Ile Glu Ile Pro Phe Lys Phe His Met Leu His Ser Gly Leu
141         370                     375                     380
142 Val His Gly Leu Ala Phe Trp Phe Asp Val Ala Phe Ile Gly Ser Ile
143 385                     390                     395                 400
144 Met Thr Val Trp Leu Ser Thr Ala Pro Thr Glu Pro Leu Thr His Trp
145             405                     410                     415
146 Tyr Gln Val Arg Cys Leu Phe Gln Ser Pro Leu Phe Ala Lys Ala Gly
147             420                     425                     430
148 Asp Thr Leu Ser Gly Thr Cys Leu Leu Ile Ala Asn Lys Arg Gln Ser
149             435                     440                     445
150 Tyr Asp Ile Ser Ile Val Ala Gln Val Asp Gln Thr Gly Ser Lys Ser
151         450                     455                     460
152 Ser Asn Leu Leu Asp Leu Lys Asn Pro Phe Phe Arg Tyr Thr Gly Thr
153 465                     470                     475                 480
154 Thr Pro Ser Pro Pro Pro Gly Ser His Tyr Thr Ser Pro Ser Glu Asn
155             485                     490                     495
156 Met Trp Asn Thr Gly Ser Thr Tyr Asn Leu Ser Ser Gly Val Ala Val
157             500                     505                     510
158 Ala Gly Met Pro Thr Ala Tyr Asp Leu Ser Ser Val Ile Ala Gly Gly
159             515                     520                     525
```

```
160 Ser Ser Val Gly His Asn Asn Leu Ile Pro Leu Ala Asn Thr Gly Ile
161     530                     535                 540
162 Val Asn His Thr His Ser Arg Met Gly Ser Ile Met Ser Thr Gly Ile
163 545                 550                 555                 560
164 Val Gln Gly Ser Ser Gly Ala Gln Gly Gly Gly Gly Ser Ser Ser Ala
165                 565                 570                 575
166 His Tyr Ala Val Asn Asn Gln Phe Thr Met Gly Gly Pro Ala Ile Ser
167             580                 585                 590
168 Met Ala Ser Pro Met Ser Ile Pro Thr Asn Thr Met His Tyr Gly Ser
169         595                 600                 605
171 <210> SEQ ID NO: 3
172 <211> LENGTH: 608
173 <212> TYPE: PRT
174 <213> ORGANISM: Artificial Sequence
176 <220> FEATURE:
177 <223> OTHER INFORMATION: CARM1 VLD TO AAA Variant
179 <400> SEQUENCE: 3
180 Met Ala Ala Ala Ala Ala Thr Ala Val Gly Pro Gly Ala Gly Ser Ala
181  1                   5                  10                  15
182 Gly Val Ala Gly Pro Gly Gly Ala Gly Pro Cys Ala Thr Val Ser Val
183             20                  25                  30
184 Phe Pro Gly Ala Arg Leu Leu Thr Ile Gly Asp Ala Asn Gly Glu Ile
185         35                  40                  45
186 Gln Arg His Ala Glu Gln Gln Ala Leu Arg Leu Glu Val Arg Ala Gly
187     50                  55                  60
188 Pro Asp Ala Ala Gly Ile Ala Leu Tyr Ser His Glu Asp Val Cys Val
189 65                  70                  75                  80
190 Phe Lys Cys Ser Val Ser Arg Glu Thr Glu Cys Ser Arg Val Gly Arg
191             85                  90                  95
192 Gln Ser Phe Ile Ile Thr Leu Gly Cys Asn Ser Val Leu Ile Gln Phe
193             100                 105                 110
194 Ala Thr Pro His Asp Phe Cys Ser Phe Tyr Asn Ile Leu Lys Thr Cys
195     115                 120                 125
196 Arg Gly His Thr Leu Glu Arg Ser Val Phe Ser Glu Arg Thr Glu Glu
197     130                 135                 140
198 Ser Ser Ala Val Gln Tyr Phe Gln Phe Tyr Gly Tyr Leu Ser Gln Gln
199 145                 150                 155                 160
200 Gln Asn Met Met Gln Asp Tyr Val Arg Thr Gly Thr Tyr Gln Arg Ala
201             165                 170                 175
202 Ile Leu Gln Asn His Thr Asp Phe Lys Asp Lys Ile Ala Ala Ala Val
203             180                 185                 190
204 Gly Cys Gly Ser Gly Ile Leu Ser Phe Phe Ala Ala Gln Ala Gly Ala
205         195                 200                 205
206 Arg Lys Ile Tyr Ala Val Glu Ala Ser Thr Met Ala Gln His Ala Glu
207     210                 215                 220
208 Val Leu Val Lys Ser Asn Asn Leu Thr Asp Arg Ile Val Val Ile Pro
209 225                 230                 235                 240
210 Gly Lys Val Glu Glu Val Ser Leu Pro Glu Gln Val Asp Ile Ile Ile
211             245                 250                 255
```

RAW SEQUENCE LISTING   DATE: 12/07/2000
PATENT APPLICATION: US/09/464,377   TIME: 10:03:40

Input Set : A:\726sl.txt
Output Set: N:\CRF3\12072000\I464377.raw

```
212 Ser Glu Pro Met Gly Tyr Met Leu Phe Asn Glu Arg Met Leu Glu Ser
213             260                 265                 270
214 Tyr Leu His Ala Lys Lys Tyr Leu Lys Pro Ser Gly Asn Met Phe Pro
215         275                 280                 285
216 Thr Ile Gly Asp Val His Leu Ala Pro Phe Thr Asp Glu Gln Leu Tyr
217     290                 295                 300
218 Met Glu Gln Phe Thr Lys Ala Asn Phe Arg Tyr Gln Pro Ser Phe His
219 305             310                 315                 320
220 Gly Val Asp Leu Ser Ala Leu Arg Gly Ala Ala Val Asp Glu Tyr Phe
221             325                 330                 335
222 Arg Gln Pro Val Val Asp Thr Phe Asp Ile Arg Ile Leu Met Ala Lys
223         340                 345                 350
224 Ser Val Lys Tyr Thr Val Asn Phe Leu Glu Ala Lys Glu Gly Asp Leu
225         355                 360                 365
226 His Arg Ile Glu Ile Pro Phe Lys Phe His Met Leu His Ser Gly Leu
227     370                 375                 380
228 Val His Gly Leu Ala Phe Trp Phe Asp Val Ala Phe Ile Gly Ser Ile
229 385             390                 395                 400
230 Met Thr Val Trp Leu Ser Thr Ala Pro Thr Glu Pro Leu Thr His Trp
231             405                 410                 415
232 Tyr Gln Val Arg Cys Leu Phe Gln Ser Pro Leu Phe Ala Lys Ala Gly
233             420                 425                 430
234 Asp Thr Leu Ser Gly Thr Cys Leu Leu Ile Ala Asn Lys Arg Gln Ser
235         435                 440                 445
236 Tyr Asp Ile Ser Ile Val Ala Gln Val Asp Gln Thr Gly Ser Lys Ser
237     450                 455                 460
238 Ser Asn Leu Leu Asp Leu Lys Asn Pro Phe Phe Arg Tyr Thr Gly Thr
239 465             470                 475                 480
240 Thr Pro Ser Pro Pro Pro Gly Ser His Tyr Thr Ser Pro Ser Glu Asn
241             485                 490                 495
242 Met Trp Asn Thr Gly Ser Thr Tyr Asn Leu Ser Ser Gly Val Ala Val
243             500                 505                 510
244 Ala Gly Met Pro Thr Ala Tyr Asp Leu Ser Ser Val Ile Ala Gly Gly
245         515                 520                 525
246 Ser Ser Val Gly His Asn Asn Leu Ile Pro Leu Ala Asn Thr Gly Ile
247     530                 535                 540
248 Val Asn His Thr His Ser Arg Met Gly Ser Ile Met Ser Thr Gly Ile
249 545             550                 555                 560
250 Val Gln Gly Ser Ser Gly Ala Gln Gly Gly Gly Ser Ser Ser Ala
251             565                 570                 575
252 His Tyr Ala Val Asn Asn Gln Phe Thr Met Gly Gly Pro Ala Ile Ser
253         580                 585                 590
254 Met Ala Ser Pro Met Ser Ile Pro Thr Asn Thr Met His Tyr Gly Ser
255     595                 600                 605
257 <210> SEQ ID NO: 4
258 <211> LENGTH: 10
259 <212> TYPE: PRT
260 <213> ORGANISM: Artifical Sequence
262 <220> FEATURE:
```

VERIFICATION SUMMARY
PATENT APPLICATION: US/09/464,377

DATE: 12/07/2000
TIME: 10:03:41

Input Set : A:\726sl.txt
Output Set: N:\CRF3\12072000\I464377.raw